點型態分析：樣方(格網)分析
Point Pattern Analysis: Quadrat Analysis

https://ceiba.ntu.edu.tw/1092Geog2017_

授課教師：溫在弘
E-mail: wenthung@ntu.edu.tw

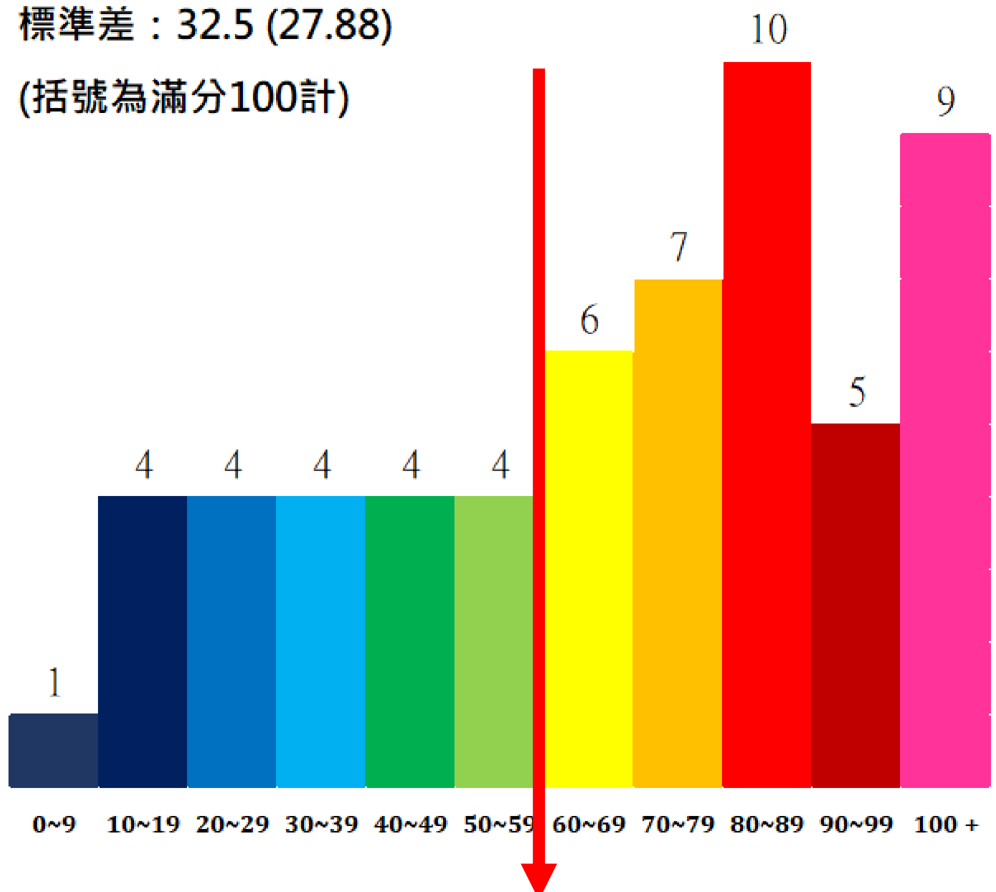
期中考成績分布

n = 58 (已扣除缺考與停修者)

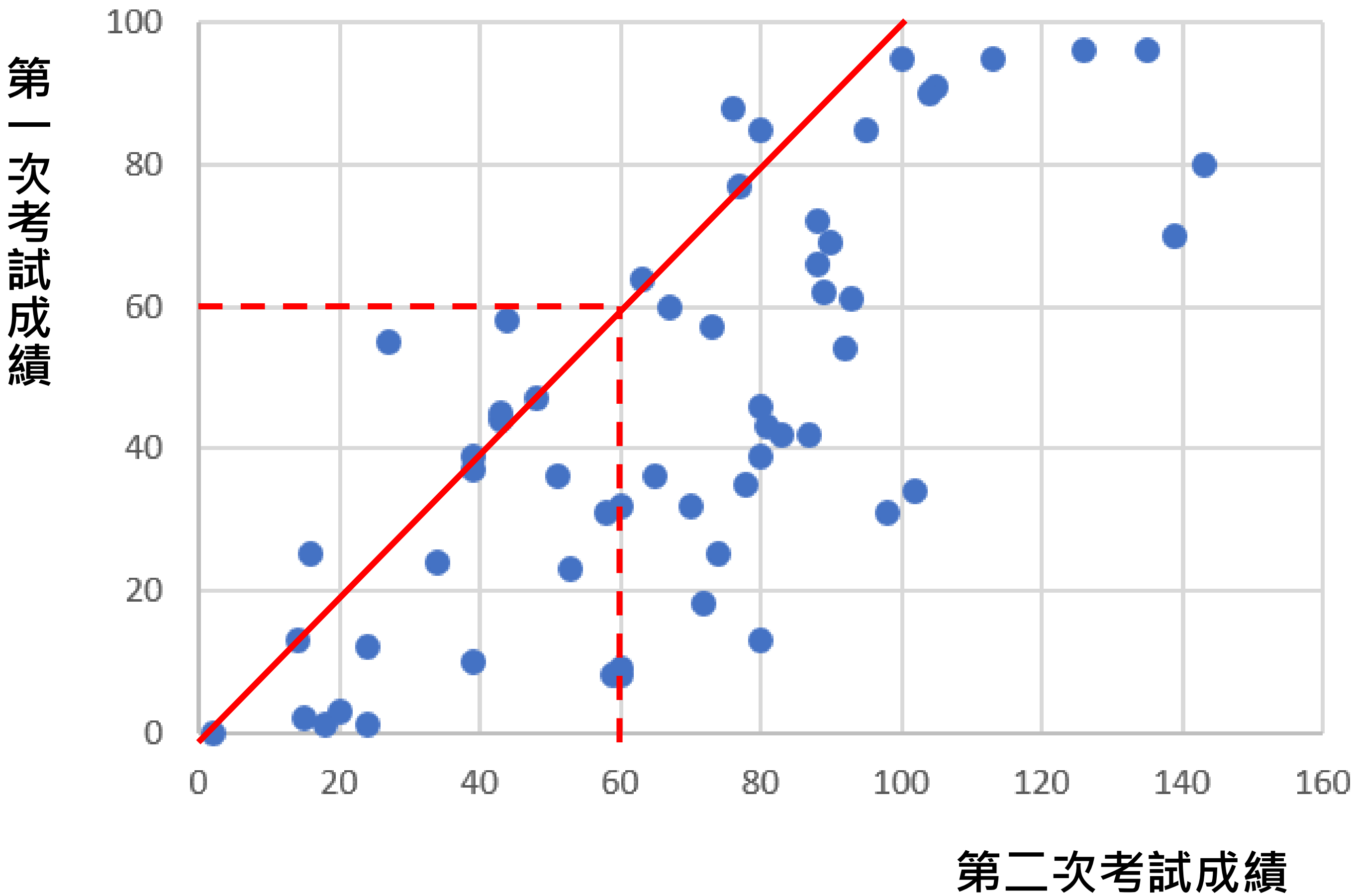
平均：68.55 (65.77)

標準差：32.5 (27.88)

(括號為滿分100計)



比較 1st vs. 2nd 期中考成績



期中考成績（30%）= Max (第1次成績, 第2次成績)

以100分為滿分計

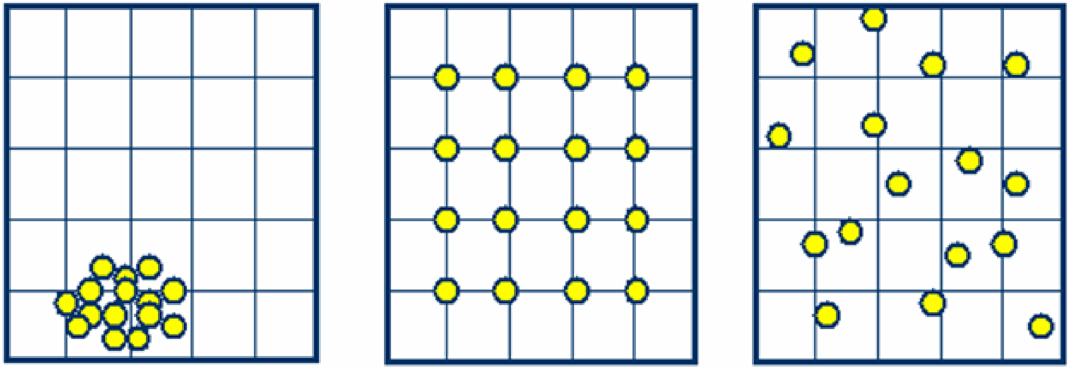
Point Pattern Analysis

Confirmatory Analysis:

Inferential Statistics

- Analyzing *Global* Patterns
  - Quadrat Analysis
  - Nearest-Neighbor Methods
  - Ripley's K Function

Global vs. Local



(a) Clustering (b) Dispersion/Uniform (c) Random

Quantitative Revolution in Geography
（50s-60s）

Environment and Planning, 1969, volume 1, pages 47-80

Quadrat analysis of urban dispersion: 1. Theoretical techniques†

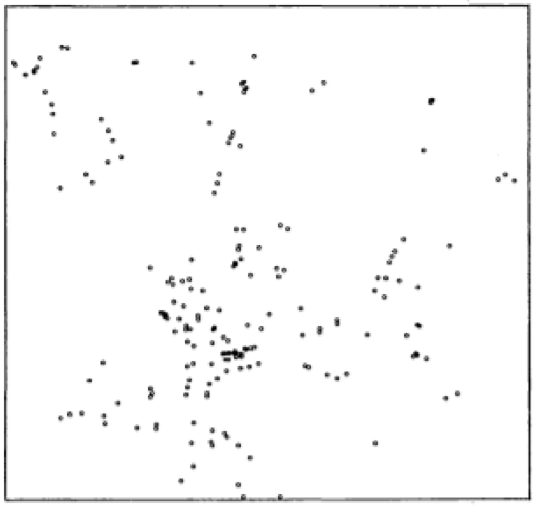
A. Rogers
Center for Planning and Development Research, University of California, Berkeley, U.S.A.
Received 10th February 1969

Quadrat analysis of urban dispersion: 2. Case studies of urban retail systems

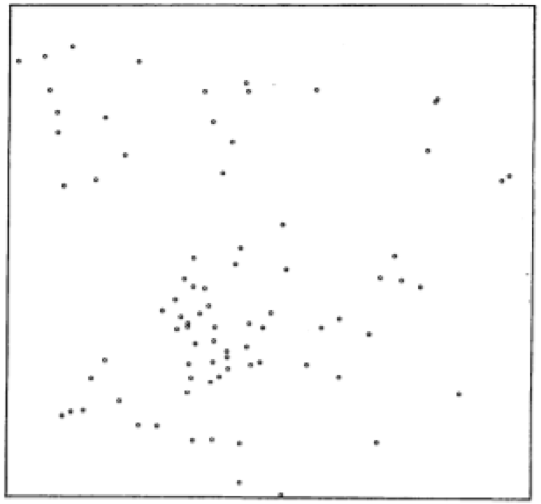
A.Rogers
Center for Planning and Development Research, University of California, Berkeley, USA
Received 10 February 1969

Abstract. This is the second part of a two part paper; the first part reviewed the methodology of quadrat analysis and in this part, two case studies are presented. A brief introduction outlines the spatial structure of retailing in urban areas and it is then demonstrated how compound and generalized distributions offer a variety of models that can be fitted to empirical data about retail spatial structure. The empirical tests use data from Ljubljana, Yugoslavia and San Francisco, California. Conclusions are drawn which relate to the description, analysis and sampling of intra-urban retail spatial dispersions.

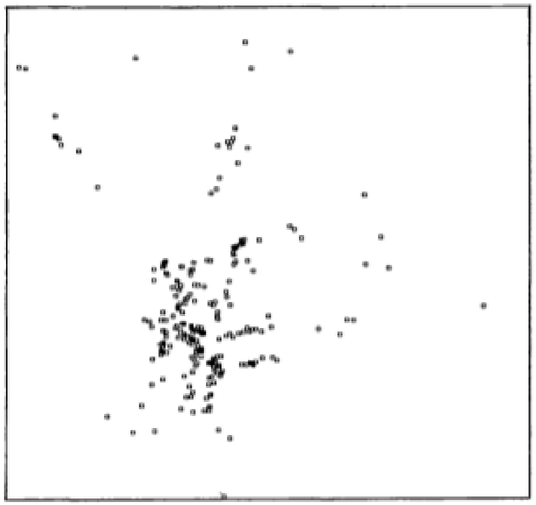
Quantitative Revolution in Geography（50s-60s）



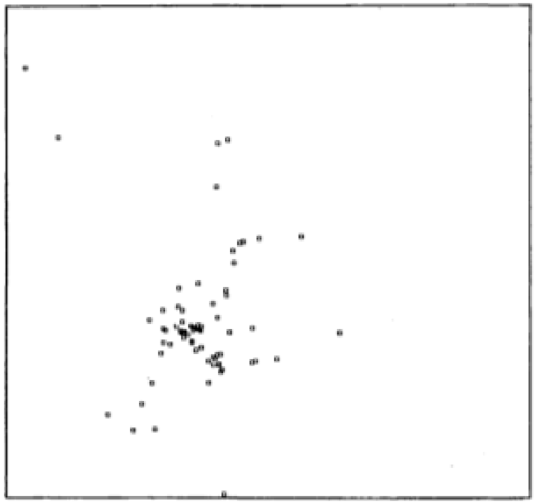
A. Food stores



B. Grocery stores



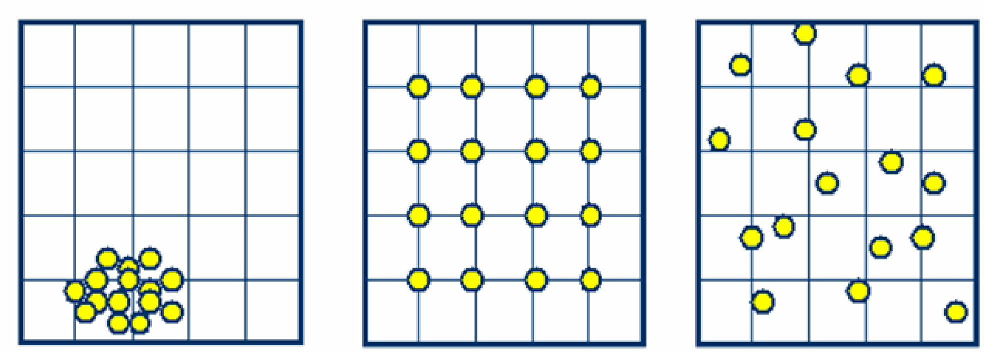
C. Non-food stores



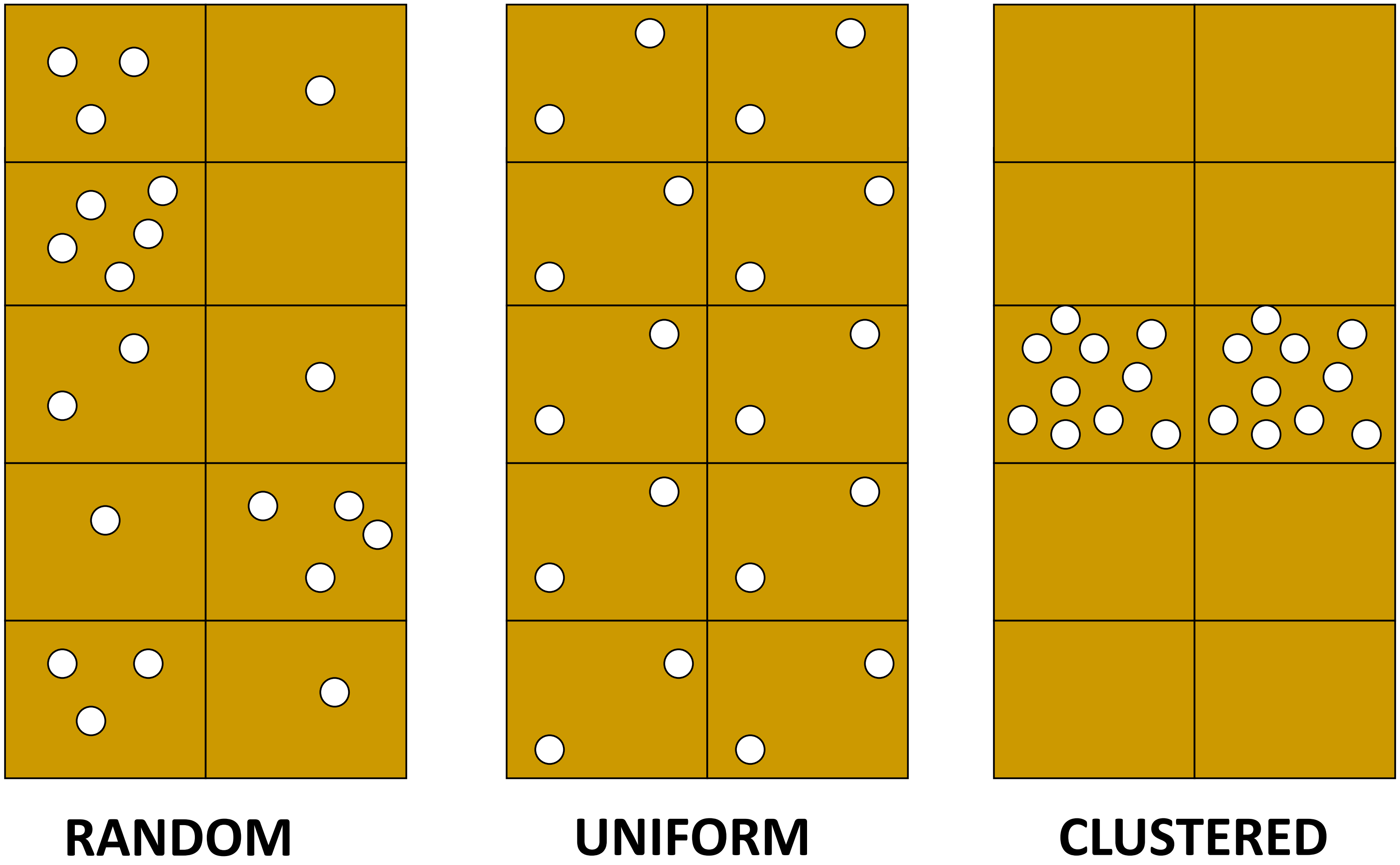
D. Clothing stores

Detection of Point Pattern

- Three general patterns
  - **Random** any point is equally likely to occur at any location and the position of any point is not affected by the position of any other point. There is no apparent ordering of the distribution
  - **Uniform** every point is as far from all of its neighbors as possible
  - **Clustered** many points are concentrated close together, and large areas that contain very few, if any, points



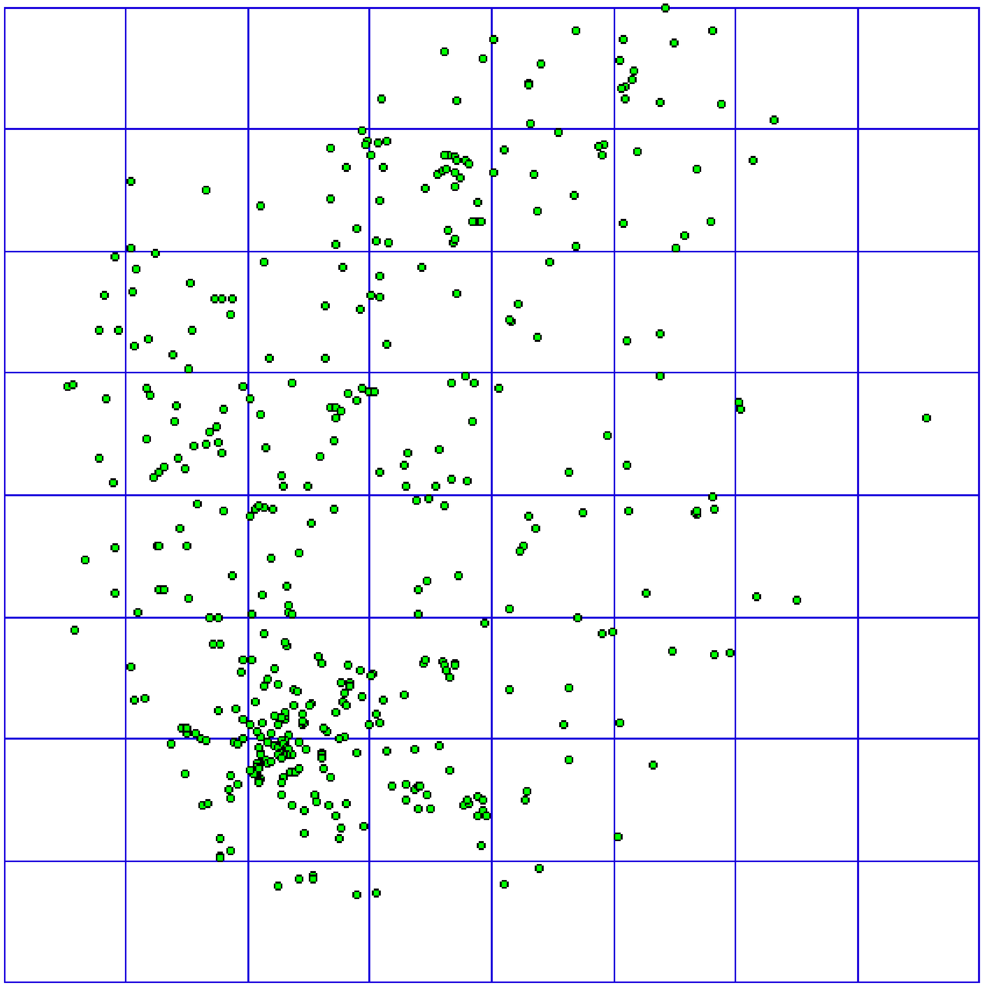
(a) Clustering (b) Dispersion/Uniform (c) Random


RANDOM
UNIFORM
CLUSTERED

Quadrat Analysis

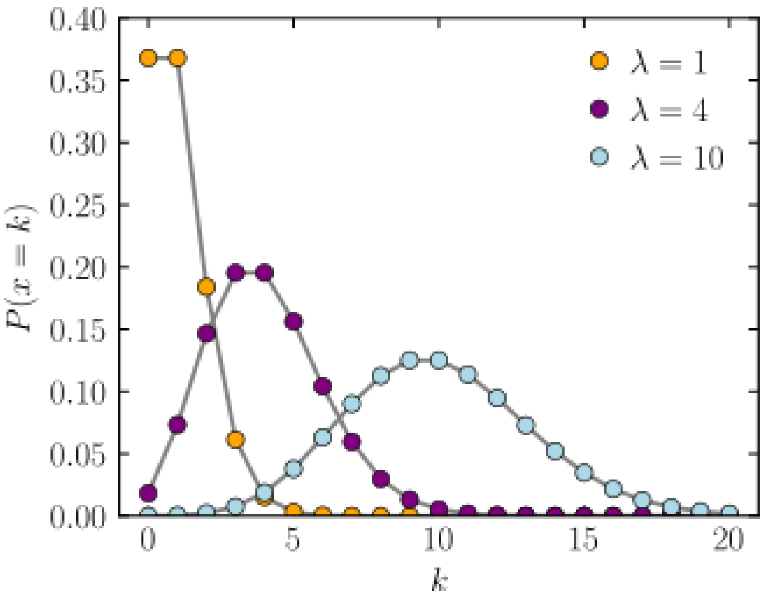
- To exam how density changes over space
- To determine if the point distribution is similar to a random distribution
  - Complete Coverage of Quadrats

Complete Coverage Quadrats



POISSON DISTRIBUTION

A probability distribution function (PDF) for discrete events



parameter $\lambda > 0$,

$$f(k;\lambda) = \Pr(X{=}k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

where

- k is the number of occurrences ($k = 0, 1, 2\ldots$)
- e is Euler's number ($e = 2.71828\ldots$)
- ! is the factorial function.

複習 Binomial distribution (二項分布)

$$P(X=k)=\binom{n}{k}p^k q^{n-k},\quad 當 k=0,1,2,\cdots,n 和 q=1-p$$

在有些應用的問題中，**n 值很大而 p 值很小**。

例如，在一個大城市裏， 我們調查一種稀有的病症。
假如每個人得病的機率都是一樣 1/500， 我們隨機的選1000人檢查，
評估這1000人中有10個人得病的機率。
應用二項分佈，我們得到

$$P(X=10)=\binom{1000}{10}\left(\frac{1}{500}\right)^{10}\left(\frac{499}{500}\right)^{990}$$

Source: http://episte.math.ntu.edu.tw/articles/mm/mm_02_4_01/index.html

複習 Binomial distribution (二項分布)

定義 $\lambda = np$

(平均數)

$$\binom{n}{k} p^k q^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!}\cdot\left(1-\frac{\lambda}{n}\right)^n \cdot \frac{n}{n}\cdot\frac{(n-1)}{n}\cdots\frac{(n-k+1)}{n}\cdot\frac{1}{(1-\frac{\lambda}{n})^k}$$

當 $n \longrightarrow \infty$ 時，如果我們維持 λ 不變（當然 p 值要趨近於零）和 k 值不變，注意

$$\frac{n-j}{n} \longrightarrow 1 \quad 當 j = 0, 1, \cdots, k-1;$$

$$\left(1-\frac{\lambda}{n}\right)^k \longrightarrow 1, \quad \left(1-\frac{\lambda}{n}\right)^n \longrightarrow e^{-\lambda}$$

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$$

所以

$$\binom{n}{k} p^k q^{n-k} \longrightarrow \frac{\lambda^k}{k!} e^{-\lambda} \qquad (2)$$

Poisson Distribution

- a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with **a known constant mean rate and independently of the time** since the last event

Properties of Poisson Distribution

A discrete random variable X is said to have a Poisson distribution with parameter $\lambda > 0$, if, for $k = 0, 1, 2, \ldots$, the probability mass function of X is given by:[6]

$$f(k;\lambda) = \Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

where

- e is Euler's number ($e = 2.71828\ldots$)
- $k!$ is the factorial of k.

The positive real number λ is equal to the expected value of X and also to its variance[7]

$$\lambda = \mathrm{E}(X) = \mathrm{Var}(X).$$

(source: Wikipedia)

[例題] 某商店每星期進進出出的客人很多（n），但客人買魚子醬的機率很小（p），只知道平均一星期賣出2罐。那麼這家商店每星期應有幾罐魚子醬的庫存？

當然不能只有2罐，因為售量超過平均數的機率很大。
當然庫存太多也會影響整個商店的運作，所以合理的庫存量是???

Using functions in R

Poisson {stats} R Documentation

The Poisson Distribution

Description

Density, distribution function, quantile function and random generation for the Poisson distribution with parameter `lambda`.

Usage

```
dpois(x, lambda, log = FALSE)
ppois(q, lambda, lower.tail = TRUE, log.p = FALSE)
qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)
rpois(n, lambda)
```

[例題] 某商店每星期進進出出的客人很多（n），但客人買魚子醬的機率很小（p），只知道平均一星期賣出2罐。那麼這家商店每星期應有幾罐魚子醬的庫存？

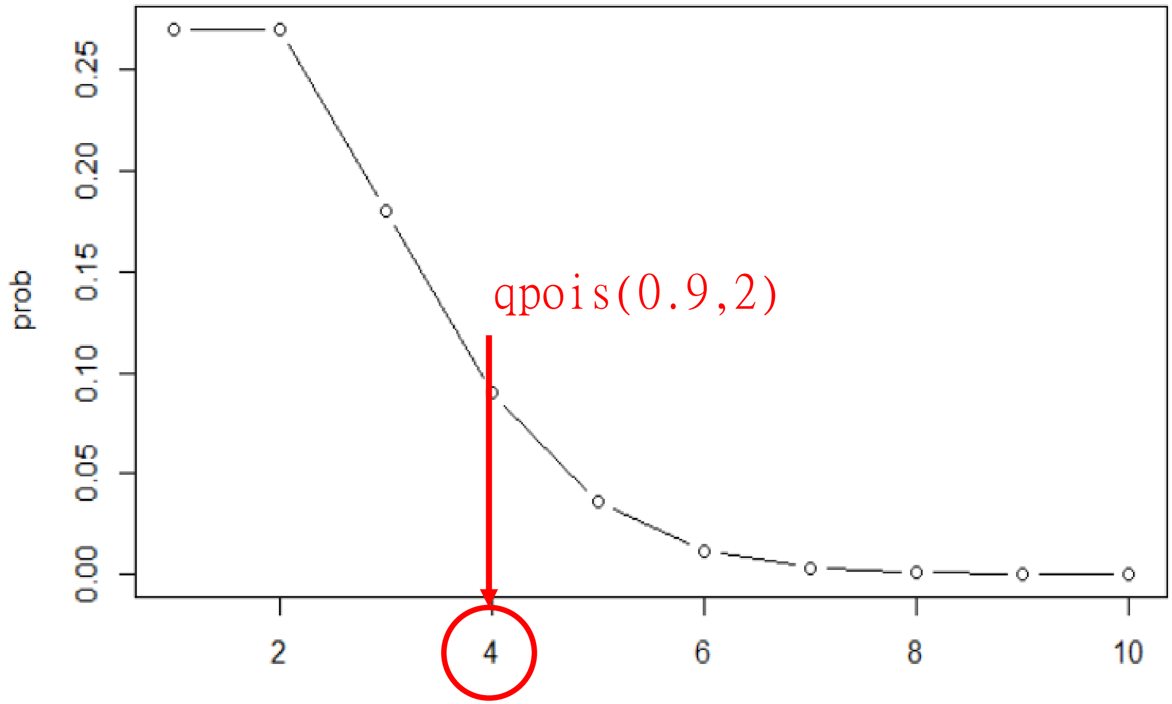
當然不能只有2罐，因為售量超過平均數的機率很大。
當然庫存太多也會影響整個商店的運作，所以合理的庫存量是???

```
prob <- vector()

for (i in 1:10){
  prob[i]<-dpois(i,2)
  }

plot(prob, type="b")

qpois(0.9,2)
```



R code: 活用 sapply( ) 取代 for-loop 迴圈

```
pdf <- sapply(1:10, function(x) dpois(x,2))

plot(pdf, type="b")
```

```
lapply(X, FUN, ...)

sapply(X, FUN, ..., simplify = TRUE, USE.NAMES = TRUE)
```

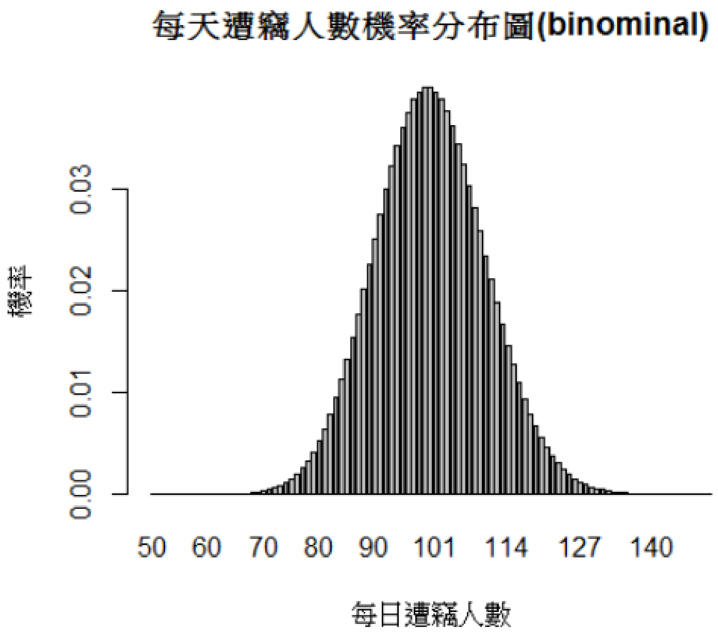
Arguments

X	a vector (atomic or list) or an expression object. Other objects (including classed objects) will be coerced by base::as.list.
FUN	the function to be applied to each element of X: see 'Details'. In the case of functions like +, %*%, the function name must be backquoted or quoted.

課堂練習： Binomial vs. Poisson Distribution

某城市有 100 萬人口， 根據過去統計分析結果，每人每天在該城市遭竊的機率 0.0001，

「每天遭竊人數」的機率分布函數 (probability distribution function, PDF)



計算隨機事件的平均值與標準差

Mean = 100

SD = 10

```
x<-rbinom(10000,size=10^6, p=0.0001)

hist(x)
mean(x)
sd(x)
var(x)
```

課堂練習： Binomial vs. Poisson Distribution

```
> dbinom(2,size = 10^6,prob = 0.000001)
[1] 0.1839398
> dpois(2, 1)
[1] 0.1839397
```

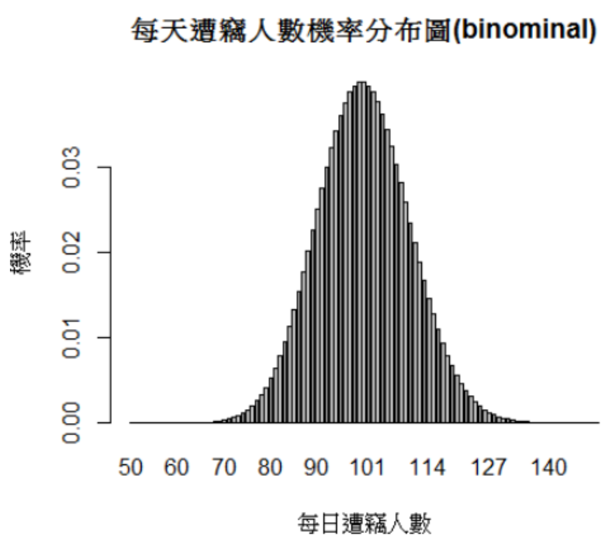
遭竊機率 = 0.0001 → 遭竊機率 = 0.000001

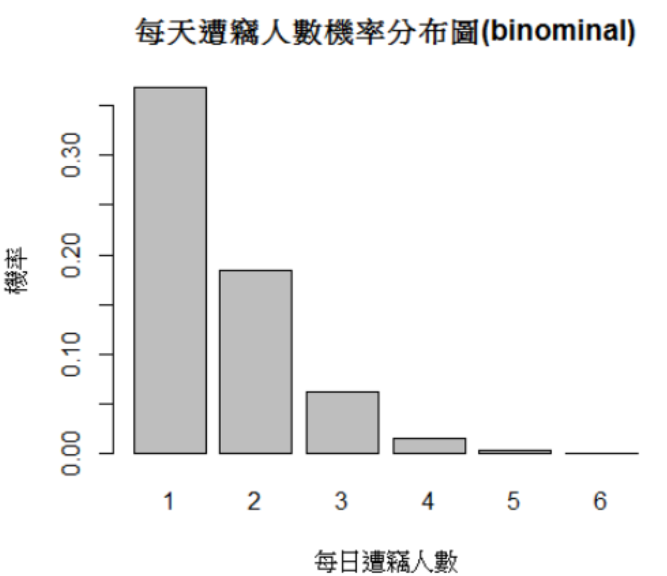
Mean = 100

SD = 10

Mean = 1.03

SD = 1.02

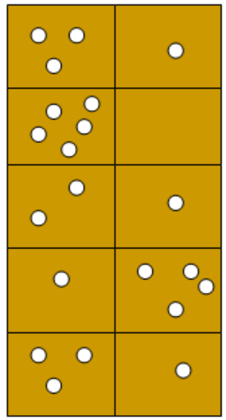




n很大時，近似常態分布 → n很大+p很小，趨近波以松分布

Quadrat Counts

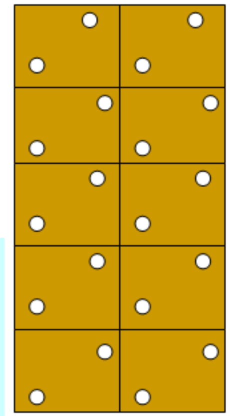
3	1
5	0
2	1
1	3
3	1



RANDOM

Quadrat #	Number of Points Per Quadrat	x^2
1	3	9
2	1	1
3	5	25
4	0	0
5	2	4
6	1	1
7	1	1
8	3	9
9	3	9
10	1	1
	20	60
Variance	2.222	
Mean	2.000	
Var/Mean	1.111	

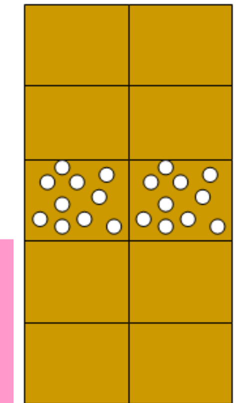
2	2
2	2
2	2
2	2
2	2



UNIFORM

Quadrat #	Number of Points Per Quadrat	x^2
1	2	4
2	2	4
3	2	4
4	2	4
5	2	4
6	2	4
7	2	4
8	2	4
9	2	4
10	2	4
	20	40
Variance	0.000	
Mean	2.000	
Var/Mean	0.000	

0	0
0	0
10	10
0	0
0	0



CLUSTERED

Quadrat #	Number of Points Per Quadrat	x^2
1	0	0
2	0	0
3	0	0
4	0	0
5	10	100
6	10	100
7	0	0
8	0	0
9	0	0
10	0	0
	20	200
Variance	17.778	
Mean	2.000	
Var/Mean	8.889	

計算公式

$$N = number\_of\_quadrats = 10$$

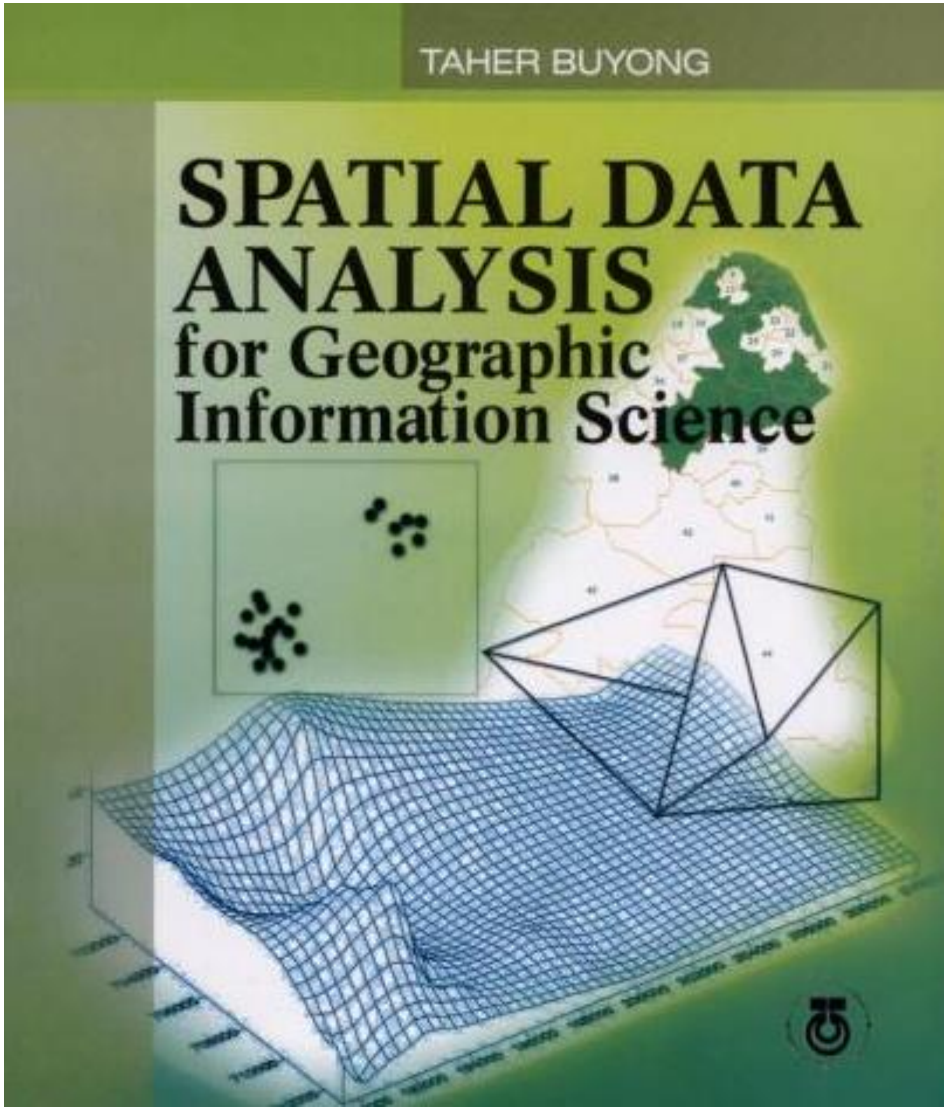
$$Variance = \frac{\sum x^2 - [(\sum x)^2 / N]}{N - 1}$$

$$Variance - mean - ratio = \frac{\text{variance}}{\text{mean}}$$

方法 1：Variance-Mean Ratio Test

- Because the mean and variance of ***Poisson Distribution*** are both equal to λ
- So the ***Variance-Mean Ratio*** should be 1.0
  - Compute the difference of the (V/M -1.0)
  - Standardize the difference by the standard error
  - Compare the Standardized score of difference to the Critical value (1.96 at 0.05 level of significance)

計算範例介紹



Source: Taher Buyong (2007), **Spatial Data Analysis for Geographic Information Science.** Penerbit UTM, ISBN 9835204233

No. of Points: 21

No. of Grids: 100

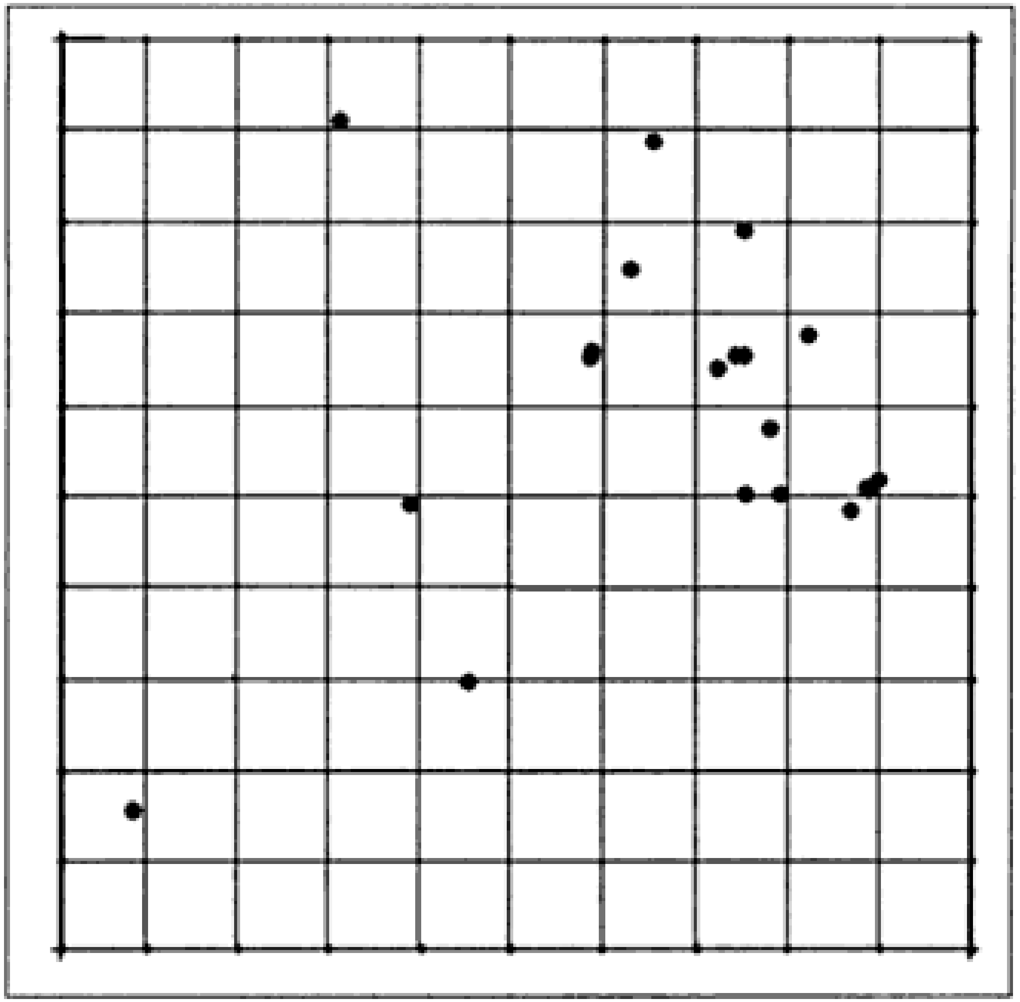


Figure 3.12 Dengue incidences in the study area with quadrats drawn

Source: Taher Buyong (2007), ***Spatial Data Analysis for Geographic Information Science*****.** Penerbit UTM, ISBN 9835204233

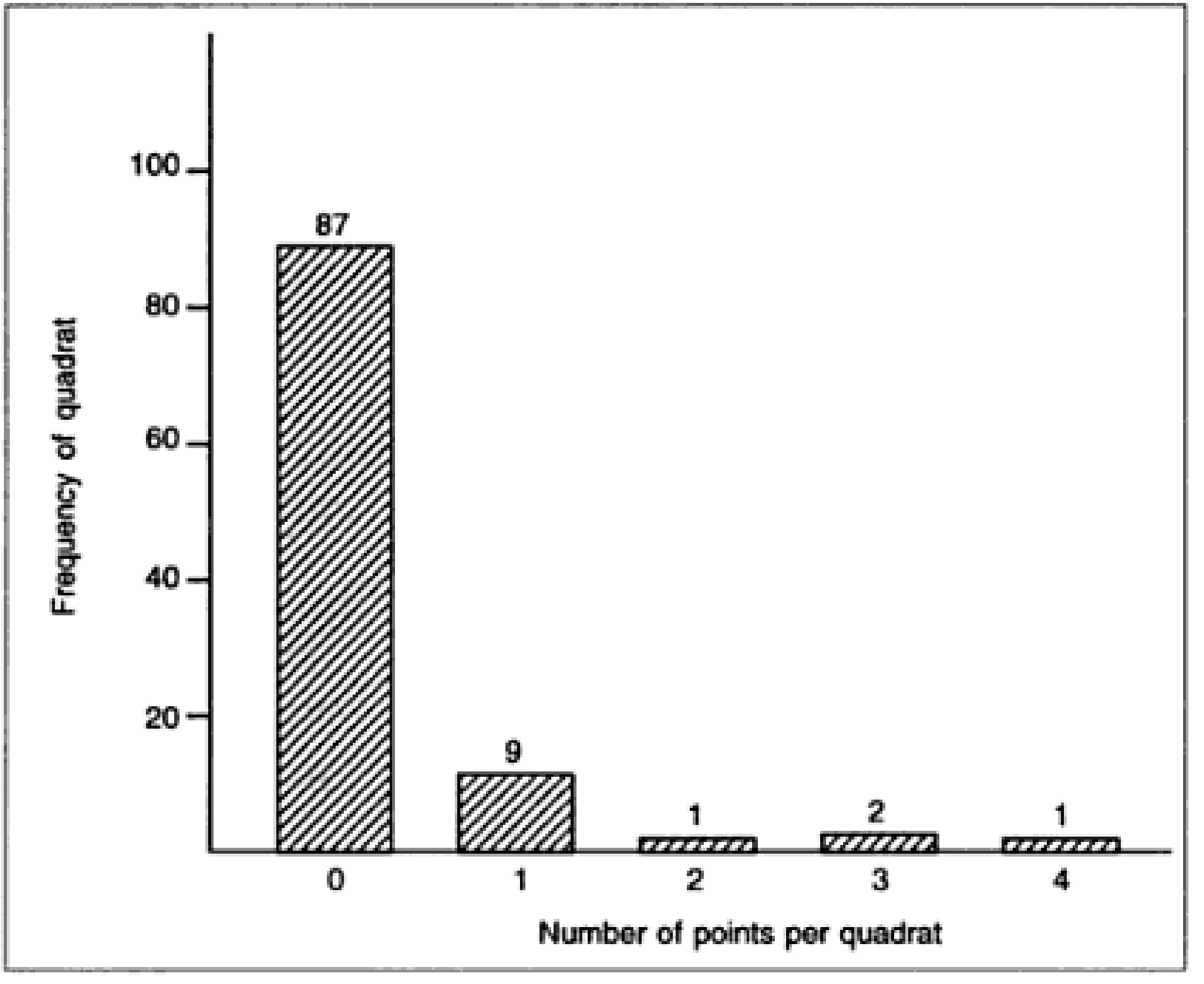


Figure 3.13 Frequency histogram of the dengue incidences

Source: Taher Buyong (2007), ***Spatial Data Analysis for Geographic Information Science***.
Penerbit UTM, ISBN 9835204233

$H_0 : s^2 = \lambda$ (pattern is random).
$H_1 : s^2 \neq \lambda$ (pattern is not random).

The test is carried out at 95% confidence level. Table 3.5 shows the necessary calculations for the test.

Table 3.5 Necessary calculations for the variance/mean ratio test

Number of points (m_i)	Frequency of quadrats (f_i)	($f_i m_i$)	($m_i - \lambda$)	$f_i (m_i - \lambda)^2$
0	87	0	−1.21	3.837
1	9	9	0.79	5.617
2	1	2	1.79	3.204
3	2	6	2.79	15.568
4	1	4	3.79	14.364
	Σ = 100	Σ = 21		Σ = 42.590

Source: Taher Buyong (2007), ***Spatial Data Analysis for Geographic Information Science***. Penerbit UTM, ISBN 9835204233

The mean:

$$\lambda = \frac{\sum f_i m_i}{\sum f_i} = \frac{21}{100}$$
$$= 0.21$$

The variance:

$$s^2 = \frac{\sum f_i (m_i - \bar{m})^2}{\sum f_i - 1} = \frac{42.59}{99}$$
$$= 0.43$$

The variance/mean ratio:

$$vmr = \frac{s^2}{\lambda} = \frac{0.43}{0.21}$$
$$= 2.05$$

The standard error of the *vmr*:

$$s_{vmr} = \left(\frac{2}{k-1}\right)^{1/2} = \left(\frac{2}{100-1}\right)^{1/2}$$
$$= 0.14$$

Source: Taher Buyong (2007), ***Spatial Data Analysis for Geographic Information Science***. Penerbit UTM, ISBN 9835204233

The test statistic:

$$t = \frac{vmr - 1}{s_{vmr}} = \frac{2.05 - 1}{0.14}$$
$$= 7.38$$

The critical values obtained from the t distribution for $\alpha = 0.025$ (two-tailed test), and $\nu = 99$ are -1.96 and 1.96. Therefore, the test fails and H_0 is rejected. Alternatively, H_1 which states that point pattern is not random is accepted. When the statistical testing is redo with $H_1 : s^2 > \lambda$ (pattern is clustered), the critical value from the t distribution for $\alpha = 0.05$ (one-tailed test) is 1.65. The test also fails and H_0 is rejected, and H_1 which states that point pattern is clustered is accepted.

Source: Taher Buyong (2007), ***Spatial Data Analysis for Geographic Information Science***. Penerbit UTM, ISBN 9835204233

方法 2：Chi-square test 卡方檢定

$$X^2 = \sum_{i=1}^{k} \frac{(x_i - m_i)^2}{m_i}$$

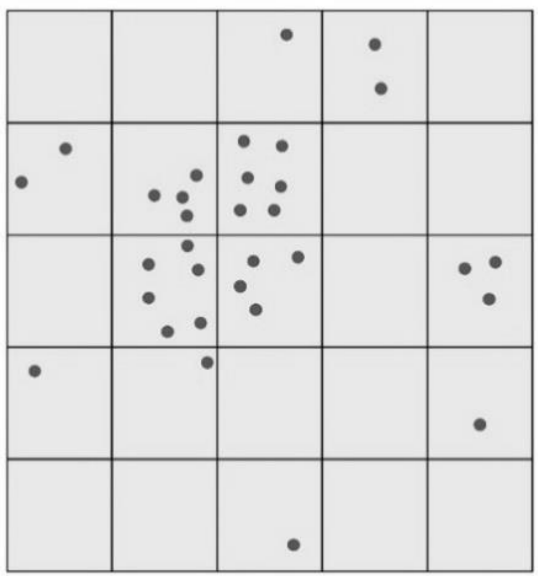
實際觀察次數 $x_i \ (i = 1, 2, \ldots, k)$

理論期望次數 $m_i = np_i$

有些數值太小，故不適用卡方檢定

格子(n)：25
點個數(x)：32

Mean = 32/25
= 1.28

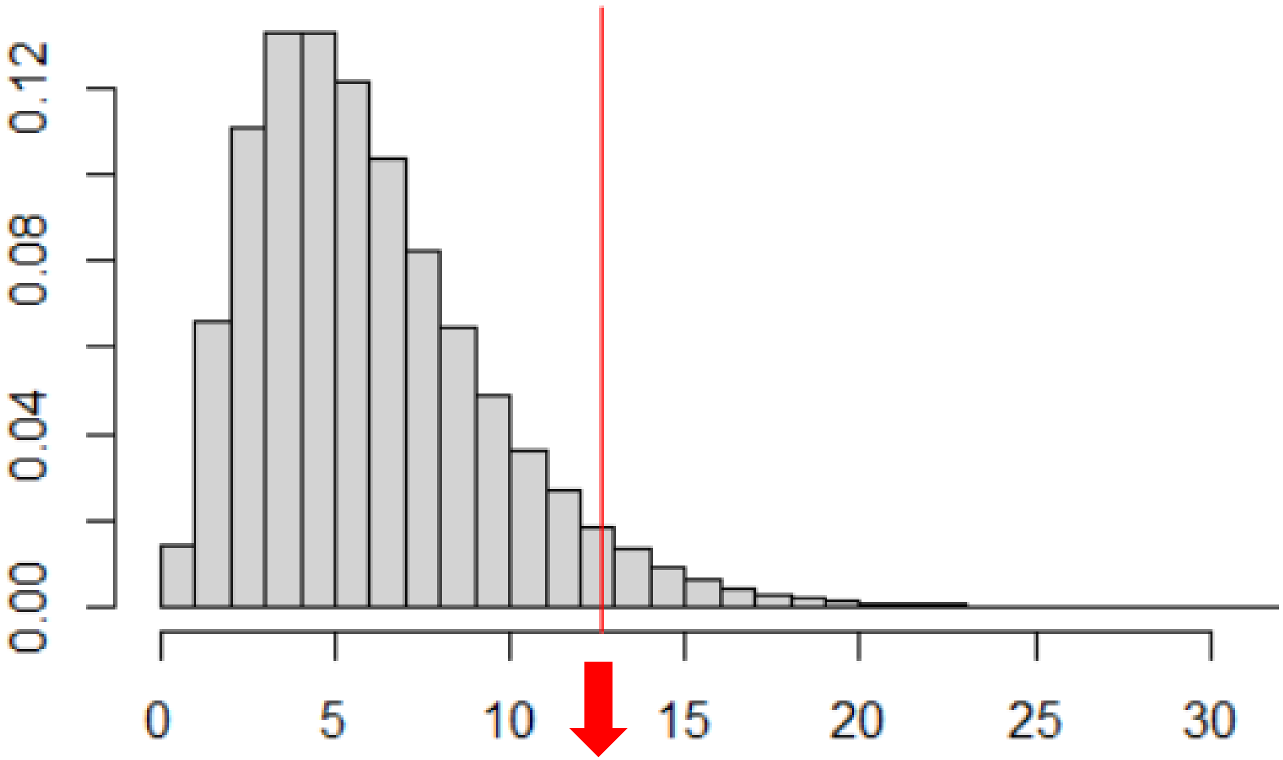


地物數量	觀察值	期望值	差值	差值平方
0	13	6.95	6.05	36.59
1	5	8.90	– 3.90	15.19
2	2	5.69	– 3.69	13.65
3	1	2.43	– 1.43	2.04
4	2	0.78	1.22	1.49
5	0	0.20	– 0.20	0.40
6	2	0.04	1.96	3.83
總和	25	24.99	0.01	73.19

x = 0 的期望值：dpois(0,1.28) x 25 = 6.95

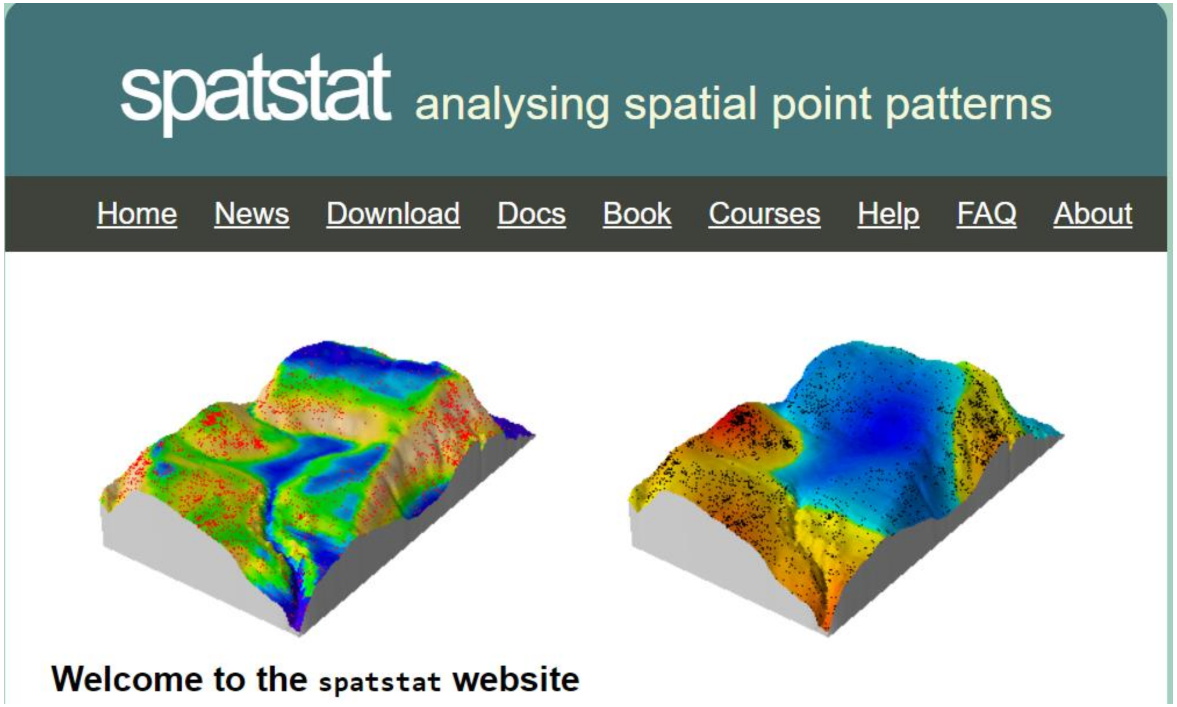
Chi-square test

The Chi-Squared prob. density distribution with 6 degrees of freedom



qchisq(0.95, df=6) = 12.59

補充 R code: Using **quadrat.test()**



spatstat (version 1.64-1)

quadrat.test: Dispersion Test for Spatial Point Pattern Based on Quadrat Counts

Description

Performs a test of Complete Spatial Randomness for a given point pattern, based on quadrat counts. Alternatively performs a goodness-of-fit test of a fitted inhomogeneous Poisson model. By default performs chi-squared tests; can also perform Monte Carlo based tests.

MS Windows 視窗作業系統之前的電腦軟體

STATISTICS IN MEDICINE, VOL. 15, 939–941 (1996)

QUADRAT ANALYSIS SOFTWARE FOR THE DETECTION OF SPATIAL OR TEMPORAL CLUSTERING

ADAM G. SKELTON

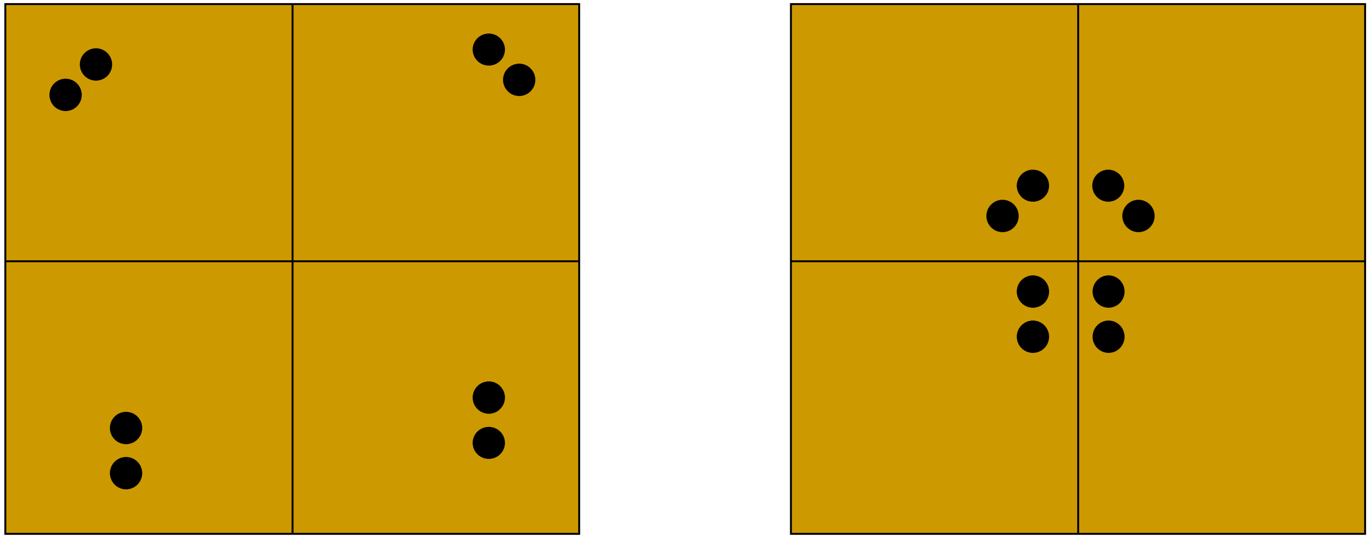
St. Louis University School of Public Health and Community Healthcare Network Modeling, 4349 Westminster Pl. St. Louis, MO 63108, U.S.A.

SUMMARY

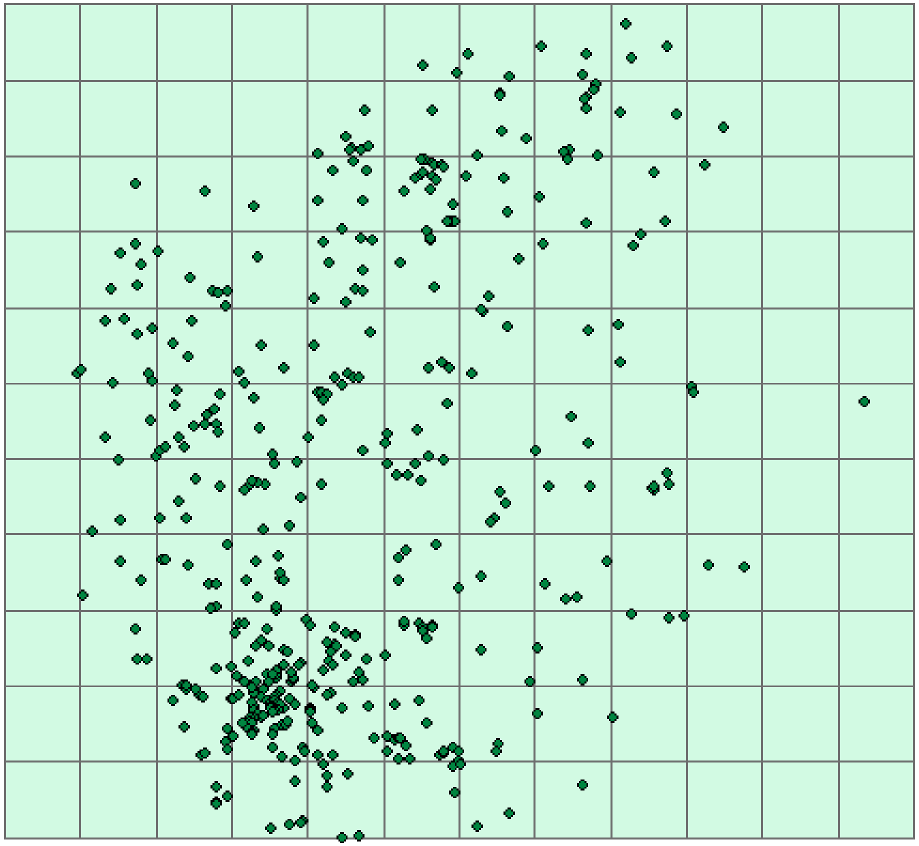
A software program is presented which automates the quadrat analysis procedure for detecting spatial or temporal clustering outlined by Boots and Getis.[1] The software essentially compares observed values within areal or temporal sampling units to those values that would have been obtained were the values distributed randomly. A chi-square test of significance is used to assess whether the values are dispersed in a non-random pattern and a *t*-test is used to assess whether the values are dispersed in a non-regular pattern. The software requires DOS 3.3 or higher and reads dBase-compatable data files.

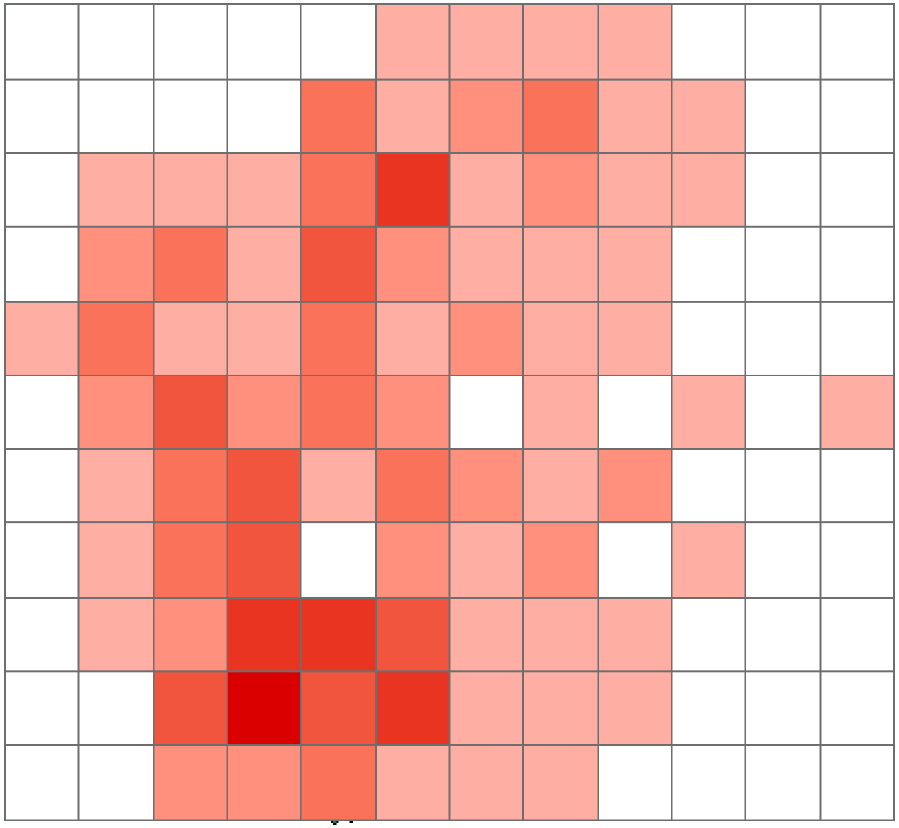
Weakness of Quadrat Method

- Results in a single measure for the entire distribution, so variations within the region are not recognized



本週實習：台南市學校的空間型態檢定





實習的操作程序

- Step 1: fishnet: st_make_grid()
- Step 2: spatial intersection: st_intersection()
- Step 3: calculate counts of points: summarise()
- Step 4: calculate mean and variance of counts
- Step 5: hypothesis testing (Variance-Mean Ratio Test)
- Step 6: make a conclusion

本週作業：台灣祭祀圈的空間型態分析

背景說明

- 「祭祀圈」係指"**<u>一個以主祭神為中心，共同舉行祭祀活動的信徒所屬的地域單位</u>**"。而空間層級由小至大可分為聚落型祭祀圈、村落型祭祀圈、全鎮型祭祀圈。根據瞿海源(1999)的台灣社會調查結果，台灣人平均有31.7%的人口會參與民間信仰的進香活動，因此對於祭祀圈的空間分析，對於瞭解台灣民間信仰的地理特性具有相當重要的角色。
- **本次實習對於「祭祀圈」的定義：祭祀圈以主祭神所在的廟宇為中心，而「村落型祭祀圈」的影響半徑約為2公里**。若全台舉辦盛大媽祖繞境活動，並且各地信眾均從社會網絡較為密切的「村落型祭祀圈」之中心廟宇出發，請利用提供的資料 (temples.rar)，回答下列問題：

寺廟圖資的資料來源：台灣寺廟網
http://www.twgod.com/

作業題目：台灣祭祀圈的空間型態分析

以下題目以台灣本島為分析範圍；

1. 若所有信仰媽祖的信眾皆從所屬祭祀圈之中心廟宇出發，要決定其中一間寺廟作為各地遶境團體匯流的折返點，如何使所有信眾的直線距離總和最小化？
2. 承上題，若主辦單位決定改成另外搭設遶境大會中心，作為匯流折返點，同樣欲使所有信眾的起終直線距離總和最小化，其中心設置的位置應於何處？(鄉鎮名稱)

以下題目以南部地區(嘉義、台南、高雄與屏東縣市)為分析範圍；

1. 比較對於信仰「媽祖」與「觀音菩薩」信眾的空間分佈差異（比較的項目，應包括：幾何中心、標準距離及標準橢圓）
2. 利用Quadrat analysis，以20 km x 20 km的網格，比較信仰「媽祖」與「觀音菩薩」的寺廟的空間群聚特性，並討論之。
3. 利用Quadrat analysis，比較20 km x 20 km vs. 50 km x 50 km的網格，計算信仰「媽祖」寺廟的空間群聚特性的差異，並討論網格尺度對於檢定空間群聚之影響。